**REPLY**

# Reply to the comments by Vorland et al. on our paper: "low-phytate wholegrain bread instead of high-phytate wholegrain bread in a total diet context did not improve iron status of healthy Swedish females: a 12-week, randomized, parallel-design intervention study"

Michael Hoppe[1,2] · Alastair B. Ross[3,4] · Cecilia Svelander[4,5] · Ann-Sofie Sandberg[4] · Lena Hulthén[2]

We thank Vorland et al. for their interest in our article and helpful comments relating to our statistical analysis and reporting. We have now reported the non-significant $p$ values in Table 1 and agree that these will be useful for powering future studies in the area of phytate reduction and iron status. We acknowledge that we have discussed within-group differences more than between-group differences, mainly because the results were counter to our hypothesis that lowering the dietary phytate should lead to improved iron status and thus required further explanation. This is also related to the between-group findings, as it would have been expected that iron status would improve with reduced phytate. During re-analysis, we have detected a between-group difference for total body iron, which supports our overall conclusion that the low-phytate bread reduced iron status, most likely due to the strong acidic flavour which altered dietary habits.

Vorland et al. suggested that we use Bonferroni correction for the $p$ values. We agree that the correction for multiple testing should be advised if there are no preplanned hypotheses and a large number of tests are carried out (Streiner

DL, Norman GR. Correction for multiple testing: is there a resolution? Chest. 2011 Jul;140;1:16–18). There is debate about the general application of Bonferroni adjustment for $p$ values, with some epidemiologists suggesting that they are, at best, unnecessary and, at worst, deleterious to sound statistical inference (Nakagawa S. A farewell to Bonferroni: the problems of low statistical power and publication bias. Behavioral Ecology. 2004;15;6:1044–1045. Perneger TV. What's wrong with Bonferroni adjustments. Brit Med J.1998;316:1236–8. Savitz DA, Olshan AF. Multiple comparisons and related issues in the interpretation of epidemiologic data. Am J Epidemiol. 1995;142:904–908. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology. 1990;1:43–6). Bonferroni correction is appropriate for e.g. genomic experiments where there are thousands of variables and the study is focused on generating a hypothesis, and the risk for Type 1 (false positive) errors is very high. However, in intervention studies that have been designed to test a specific hypothesis, such as ours (bread with low phytate increases iron status compared to bread with natural levels of phytate), use of Bonferroni or other false discovery rate correction factors is not necessary. We have measured eight variables (Table 1). Five of the

This Reply refers to the letter to the editor published here: https://doi.org/10.1007/s00394-020-02287-0

The original publication can be read here: https://link.springer.com/article/10.1007/s00394-018-1722-1.

✉ Michael Hoppe
michael.hoppe@nutrition.gu.se

Alastair B. Ross
Alastair.Ross@agresearch.co.nz

Cecilia Svelander
cecilia.svelander@regionorebrolan.se

Ann-Sofie Sandberg
ann-sofie.sandberg@chalmers.se

Lena Hulthén
lena.hulthen@medfak.gu.se

1    Department of Gastroenterology and Hepatology, Clinical Nutrition Unit, Sahlgrenska University Hospital, Göteborg, Sweden

2    Department of Internal Medicine and Clinical Nutrition, Sahlgrenska Academy at the University of Gothenburg, Göteborg, Sweden

3    Lincoln Research Centre, AgResearch, Springs Road, Lincoln, Christchurch, New Zealand

4    Department of Biology and Biological Engineering, Food and Nutrition Science, Chalmers University of Technology, Göteborg, Sweden

5    Örebro University Hospital, Örebro, Sweden

**Table 1** Data at baseline and after 12 weeks of intervention

| | High-phytate bread group (n=31) | | | Low-phytate bread group (n=24) | | | |
|---|---|---|---|---|---|---|---|
| | Baseline | Post-intervention | p value (within group) | Baseline | Post-intervention | p value (within group) | p value (between groups) |
| BMI (kg/m²) | 22.6±0.6 | 22.8±0.6 | 0.326 | 21.7±0.5 | 21.8±0.5 | 0.145 | 0.774 |
| Hb (g/L) | 134±1 | 136±1 | 0.155 | 132±1 | 134±1 | 0.178 | 0.682 |
| Hepcidin (ng/ml) | 11.8±2.4 | 11.8±2.0 | 0.704 | 13.6±2.7 | 9.5±2.8 | 0.433 | 0.343 |
| Ferritin (µg/L) | 31±4 | 32±4 | 0.859 | 33±3 | 27±6 | **0.018** | 0.251 |
| TfR (mg/L) | 2.7±0.1 | 2.7±0.1 | 0.409 | 2.9±0.1 | 2.9±0.6 | 0.489 | 0.491 |
| Body Fe [a] (mg/kg) | 7.0±0.4 | 7.0±0.5 | 0.738 | 6.9±0.4 | 5.4±0.5 | **0.035** | **0.035** |
| Alkyl-resorcinols (mmol/L) | 351±52 | 467±77 | 0.120 | 266±68 | 522±127 | **0.002** | 0.139 |
| C17:C21 ratio | 0.16±0.02 | 0.23±0.02 | **0.018** | 0.15±0.02 | 0.29±0.03 | **0.001** | 0.051 |

Subjects were allocated into two groups that either received (on a daily basis) 200 g wholegrain rye flour-based bread natural high in phytates or 200 g dephytinized wholegrain rye flour-based bread. Evaluation was done at baseline and after 12 weeks

Values represent geometric mean ± standard error of the mean (SEM, as italicized values)

Bold values indicate the level of significance $p \leq 0.05$

[a] Calculation of body iron reserves was based on the ratio between soluble transferrin receptor and serum ferritin [1]

variables are measured at different aspects of iron status and so are not fully independent of each other. This large number of markers of iron status is necessary as there is no one single marker that adequately accounts for iron status. Although false discovery rate correction would reduce the risk for false-positive results, the balance is that we greatly increase the risk for Type 2 outcomes (false-negative results). Other standard clinical parameters are there to indicate if there was any unexpected effect on the health or body composition that may have impacted on the results, and the alkylresorcinols, biomarkers of wholegrain intake, were measured to help explain if compliance may have been an issue—also independent of the primary hypothesis, but important for understanding the study outcomes. Blanket use of a false discovery rate factor is not always appropriate and indeed rarely applied in nutrition intervention studies with a clear hypothesis and many variables measured.

When there is a high rate of dropouts, it is important to consider any potential confounding effect. We acknowledge that a common problem with the presently used approach, per-protocol (PP) analysis, is that it can result in an overestimation of the treatment effect. Thus, an alternative approach would be intent-to-treat (ITT). A problem with ITT analysis may be that there is no outcome data for patients who dropped out. Another obvious disadvantage of intention-to-treat is that the subjects are mixed up with individuals who have not received the intended intervention at all, which gives a dilution effect and a clear underestimation of the

effect in subjects who are really exposed to the intervention. Thus, in an ITT analysis, patients are not analysed according to the treatment actually received. When performing an ITT analysis of our data by using the principle of "last observation carried forward" there were no change in any of the iron status biomarkers in the high-phytate bread group (n=49). In the low-phytate bread group (n=53), there were no decrease in ferritin (p<0.053), but a decrease in total body iron (p<0.039), but no change in transferrin receptor (TfR) concentration, as reported in our paper. We have also explained the issues with the dietary intervention, and how this informed our use of PP instead of ITT. Comments from an independent statistician engaged by the European Journal of Nutrition supports our methodology for imputing missing data, although we acknowledge that there are several methods for this. Nevertheless, we have now performed between-group ITT comparisons applying the last observation carried forward approach. There were no between-group changes for markers of iron status (TfR p=0.571; Ferritin p=0.348; Body-Fe p=0.499; Hepcidin p=0.359; Using Mann–Whitney test).

In all, we thank Vorland et al. for their comments, which will help power calculations for future studies. The statistical reanalysis of our results underlines our earlier finding that counter to hypothesis, low-phytate wholegrain rye bread had a negative impact on iron status in young women. This may be related to the flavour of the test bread and the relatively high amount fed, and future studies with test products

with improved flavour and different doses may find different outcomes.

To correct the original paper [2] an erratum [3] has been published.

## References

1. Cook JD, Flowers CH, Skikne BS (2003) The quantitative assessment of body iron. Blood 101(9):3359–3364

2. Hoppe M, Ross AB, Svelander C, Sandberg AS, Hulthén L (2019) Low-phytate wholegrain bread instead of high-phytate wholegrain bread in a total diet context did not improve iron status of healthy Swedish females: a 12-week, randomized, parallel-design intervention study. Eur J Nutr 58(2):853–864. https://doi.org/10.1007/s00394-018-1722-1

3. Hoppe M, Ross AB, Svelander C, Sandberg AS, Hulthén L (2020) Correction to: Low-phytate wholegrain bread instead of high-phytate wholegrain bread in a total diet context did not improve iron status of healthy Swedish females: a 12 week, randomized, parallel-design intervention study. Eur J Nutr. https://doi.org/10.1007/s00394-020-02286-1